This problem uses a variant on the wage/education relationship using the log of wage, so that a change of 0.01 in the dependent variable is approximately an increase of 1% in the wage. Here is the simple regression of log(wage) on education:

```
. reg lwage educ

      Source |       SS       df       MS              Number of obs =    1000
-------------+------------------------------           F(  1,   998) =  216.41
       Model |  60.015841     1   60.015841            Prob > F      =  0.0000
    Residual |  276.76489   998   .27731953            R-squared     =  0.1782
-------------+------------------------------           Adj R-squared =  0.1774
       Total |  336.780731   999   .337117849          Root MSE      =  .52661


-------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+-----------------------------------------------------------------
        educ |   .0904082   .0061456    14.71   0.000     .0783484    .1024681
       _cons |   1.609444   .0864229    18.62   0.000     1.439853    1.779036
-------------------------------------------------------------------------------
```

Adding three regional dummy variables to the regression (the East region is omitted) yields

```
. reg lwage educ midwest south west

      Source |       SS       df       MS              Number of obs =    1000
-------------+------------------------------           F(  4,   995) =   58.04
       Model |  63.7111173     4  15.9277793           Prob > F      =  0.0000
    Residual |  273.069614   995   .274441823          R-squared     =  0.1892
-------------+------------------------------           Adj R-squared =  0.1859
       Total |  336.780731   999   .337117849          Root MSE      =  .52387


-------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+-----------------------------------------------------------------
        educ |   .0904304    .006124    14.77   0.000     .0784129    .1024479
     midwest |  -.1313871   .0487029    -2.70   0.007    -.2269593    -.035815
       south |  -.0523116    .046397    -1.13   0.260    -.1433588    .0387357
        west |   .0332736   .0487333     0.68   0.495    -.0623583    .1289056
       _cons |   1.648171    .092581    17.80   0.000     1.466494    1.829847
-------------------------------------------------------------------------------
```

1. Why can we not add a dummy for the East region? What would happen if we did?

2. Interpret the coefficients of the three dummy variables and their individual t statistics. (What economic hypothesis does each of these t tests test?)

3. Specify the null and alternative hypotheses for a test that, controlling for education, wages in the West are the same as in the South. Give the formula for the test statistic that you could use to test this. Why can you not calculate the test statistic from the information in the table?

4. Use the SSR form of the F test in Wooldridge's equation [4.37] to test the null hypothesis that region does not matter, $i.e.$, H_0: $\beta_2 = 0$, $\beta_3 = 0$, and $\beta_4 = 0$ against H_1: $\beta_2 \neq 0$ or $\beta_3 \neq 0$ or $\beta_4 \neq 0$ in the regression $\ln(wage) = \beta_0 + \beta_1 educ + \beta_2 midwest + \beta_3 south + \beta_4 west + u$.